



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2020

---

## **Schemas and the frequency/acceptability mismatch: Corpus distribution predicts sentence judgments**

Flach, Susanne

**Abstract:** A tight connection between competence and performance is a central tenet of the usage-based model. Methodologically, however, corpus frequency is a poor predictor of acceptability – a phenomenon known as the “frequency/acceptability mismatch”. This article argues that the mismatch arises from a “methodological mismatch”, when simple frequency measures are mapped onto complex grammatical units. To illustrate, we discuss the results of acceptability judgments of go/come-v. The construction is subject to a formal constraint (Go see the doctor! vs. \*He goes sees the doctor), which results from its mandative semantics (directives, commissives). While a formal model makes no prediction with regard to gradient acceptability of bare (“grammatical”) go/come-v, the usage-based view assumes that acceptability is a function of compatibility with an abstract schema. The experimental ratings are compared with a number of corpus-derived measures: while acceptability is largely independent of (raw) frequency, it is not independent of frequency-related usage distribution. The results add to recent suggestions that the frequency/acceptability mismatch is substantially reduced if the syntactic complexity of a unit is appropriately captured in usage data.

DOI: <https://doi.org/10.1515/cog-2020-2040>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-197117>

Journal Article

Published Version

The following work is licensed under a Publisher License.

Originally published at:

Flach, Susanne (2020). Schemas and the frequency/acceptability mismatch: Corpus distribution predicts sentence judgments. *Cognitive Linguistics*, 31(4):609-645.

DOI: <https://doi.org/10.1515/cog-2020-2040>

Susanne Flach\*

# Schemas and the frequency/acceptability mismatch: Corpus distribution predicts sentence judgments

<https://doi.org/10.1515/cog-2020-2040>

**Abstract:** A tight connection between competence and performance is a central tenet of the usage-based model. Methodologically, however, corpus frequency is a poor predictor of acceptability – a phenomenon known as the “frequency/acceptability mismatch”. This article argues that the mismatch arises from a “methodological mismatch”, when simple frequency measures are mapped onto complex grammatical units. To illustrate, we discuss the results of acceptability judgments of *go/come*-v. The construction is subject to a formal constraint (*Go see the doctor!* vs. *\*He goes sees the doctor*), which results from its mandative semantics (directives, commissives). While a formal model makes no prediction with regard to gradient acceptability of bare (“grammatical”) *go/come*-v, the usage-based view assumes that acceptability is a function of compatibility with an abstract schema. The experimental ratings are compared with a number of corpus-derived measures: while acceptability is largely independent of (raw) frequency, it is not independent of frequency-related usage distribution. The results add to recent suggestions that the frequency/acceptability mismatch is substantially reduced if the syntactic complexity of a unit is appropriately captured in usage data.

**Keywords:** frequency/acceptability mismatch, methodological mismatch, acceptability judgment task, formal constraint, constructional semantics, correspondence analysis, association, converging evidence

---

\*Corresponding author: Susanne Flach, Universität Zürich, Zürich, Switzerland; and Université de Neuchâtel, Neuchâtel, Switzerland, E-mail: [susanne.flach@es.uzh.ch](mailto:susanne.flach@es.uzh.ch).  
<https://orcid.org/0000-0002-7469-2703>

# 1 Introduction

The so-called “frequency/acceptability mismatch”, also called the “grammaticality/frequency gap”, refers to the observation that there is no reliable correlation between the frequency of a syntactic unit and its acceptability (Bader and Häussler 2010; Kempen and Harbusch 2005). While frequent patterns are fully acceptable, the prediction fails for rare or unattested sentences with “arbitrarily low probabilities” (Manning 2003: 309). The mismatch comes in two forms: in a *ceiling mismatch*, two equally acceptable units come from very different frequency bands, while two low-frequent units diverge considerably in acceptability in a *floor mismatch* (Bader and Häussler 2010: 316). The effect has been observed for a range of syntactic and morphological constructions (e.g., Arppe and Järviö 2007; Bader and Häussler 2010; Bermel and Knittl 2012; Divjak 2008; Featherston 2005; Kempen and Harbusch 2005; Manning 2003).

The mismatch is interpreted very differently between theoretical frameworks. Formal models tend to attribute it to competing modular processes that operate below a certain frequency threshold (Bader and Häussler 2010; Featherston 2005). The persistence of gradience in acceptability has also afforded suggestions that grammaticality itself may be gradient as a function of constraint accumulation (Featherston 2005; Keller 2000; Sorace and Keller 2005; Wasow 2009). Methodologically, most studies in Experimental Syntax tend to represent usage in the simplest way possible by counting how often strings of words, part-of-speech tags, or constituents occur in corpora (so-called “structural frequencies”, Bader and Häussler 2010: 313).

The mismatch is more difficult to explain for usage-based models. The assumption of an intimate relationship between performance and competence implies a close relationship of usage data and experimental behavior (Bybee 2006; Langacker 1987, 1988, 2000). The central claim is that speakers build their constructional knowledge inductively by abstracting schemas over repeated exposure to sufficiently similar instances. Since each usage event is situated in complex communicative settings, and contains formal, semantic, and pragmatic information, schemas are rich in conceptual structure, but low in specific details (Langacker 2000: 4, 10). On this view, the well-formedness of a unit depends on how well it instantiates the schema. That is, if a pattern (B) instantiates schema [A] in its full specifications, (B) is said to be grammatical (or conventional) with respect to [A]. The reverse holds for ill-formedness: the more (B) deviates from relevant specifications of [A], the greater the likelihood that (B) is unacceptable or ungrammatical. As abstractions arise bottom-up over repeated exposure, grammaticality is relative to a structure’s degree of conventionality. Each utterance is

categorized against previously established schemas (Langacker 1987, 1988, 2000) and its compatibility with an established schema varies along a number of lexical, morphological, syntactic, semantic, or social dimensions. In brief, acceptability is a gradient function of compatibility of an instance with a higher-order schema.

One advantage of the compatibility view is that it does not require the assumption of a frequency threshold, because compatibility with a schema should hold across all frequency bands. From this angle, the frequency/acceptability mismatch presents a problem: if compatibility is the result of exposure to repeated use, frequency makes the wrong predictions for exceedingly rare but fully acceptable expressions. This may lead to the assumption that compatibility with a schema is poorly represented by usage and/or frequency.

This conclusion is premature, because two things must be borne in mind. First, FREQUENCY OF USE, assumed to underlie speaker knowledge (Langacker 1987: 59), is a conceptual notion – it does not entail the empirical measure (*raw*) frequency. To say that repetition reinforces entrenchment neither means that speakers keep a counter of how often they hear a structure or string, let alone that they judge an expression based on this counter. Nor does it mean that repetition only pertains to strings or formal properties. Repetition also reinforces semantic and contextual knowledge of usage events in their social settings, leading to the rich conceptual schemas described above (Langacker 2000: 11). In other words, entrenchment is not linear, but complex, contingent, and multifactorial.

Second, schemas vary in specificity and complexity (cf. Figure 1). Simplexes are low in both schematicity and complexity (e.g., *time*, *house*). They instantiate their schema directly and are by definition compatible with it. Complexes such as semi-fixed expressions (e.g., *for NP's sake*) or syntactic structures (e.g., the ditransitive, [NP V NP NP]) are increasingly schematic: they are abstracted over higher type frequencies and complex lexico-grammatical relationships (Langacker 1987: 25–27; Stefanowitsch and Flach 2016: 106). As the schematicity of a unit increases, so does the contingency of the contextual information (e.g., type frequency, lexical associations, subpatterns, distributional skews). Therefore, units in different areas of the complexity space require different operationalizations and measurements to capture the compatibility of its instances (see Stefanowitsch and Flach 2016 for an overview).

Corpus frequency is a very good predictor of experimental behavior of monomorphemic words, which are directly compatible with their stored schema; the accuracy of prediction depends on the quality of the corpora from which the frequencies are drawn (e.g., Arnon and Snider 2010; van Heuven et al. 2014). However, the frequency/acceptability mismatch has mainly been identified for syntactic units of high complexity and schematicity. Put simply, it results from what we may call a “methodological mismatch”: if “frequency” is understood as

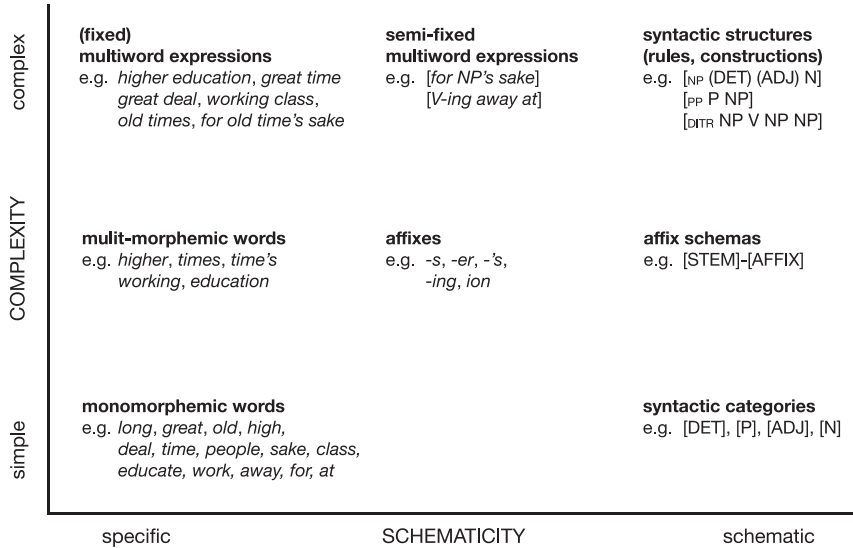


Figure 1: Linguistic units by complexity and schematicity (Stefanowitsch and Flach 2016: 106).

the number of times a unit occurs in a corpus, complex contextual properties of units further to the top right are not properly captured. This problem does not disappear if the counted *unit* is itself abstract (e.g., part of speech tags, phrase structures, constituents, or word order patterns), because their *counts* are still token frequencies or something equally one-dimensional. For instance, word order patterns are often ambiguous and can instantiate any number of distinct constructional schemas, each with their own lexical associations, distributions, and functional or structural properties: the subject-control pattern in *she put energy into completing the project* is licensed by a different schema than the object-control pattern in *she talked him into completing the project*, although both have the same linear structure (or “structural frequency”). Counting word order or constituent patterns glosses over complex properties, which is often aggravated by poor precision in automatic corpus extraction (i.e., how many hits are instances of the construction).

If, however, complex measures – e.g., verb associations, type frequencies, transitional probabilities, family sizes, distributions, skews, resemblances or relatedness – are included in a way that is suitable to capture multifactorial usage properties of the higher order schema, the frequency/acceptability mismatch is substantially reduced (Arppe and Järvikivi 2007; Bermel and Knittl 2012; Divjak 2017; Gries et al. 2005; Lau et al. 2017; Wiechmann 2008). As Divjak (2017: 372) puts it, “it is not so much the case that usage frequency has problems predicting

acceptability judgments at the low end of the frequency spectrum. It is rather the case that the wrong type of frequency data has been foregrounded". It is not surprising that converging evidence from observational and experimental data is mounting once higher-order generalizations with form–meaning correspondences (Goldberg 1995) are factored in. After all, the usage-based model assumes that approaching constructional properties this way proxies speakers' strategies in extracting schemas from repeated exposure to language in communicative settings, but not simply by keeping a (string) counter. This knowledge indeed emerges as a non-trivial influence in acceptability tasks (Divjak 2017).

The methodological mismatch (FREQUENCY OF USE  $\neq$  frequency) is probably due in no small part to the ambiguity of the term *frequency* and what we mean when we talk about frequency data. On the one hand, FREQUENCY OF USE is a foundational assumption of the usage-based model (Bybee 2006; Langacker 1987). On the other hand, the term *frequency* is often used as a short-hand for complex corpus-derived measures that operationalize the conceptual notion FREQUENCY OF USE ("usage intensity", Stefanowitsch and Flach 2016: 108). At the same time, the critique of frequency as an oversimplified measure of entrenchment pertains to its simple count-reading (e.g., Schmid 2010). Hence, a frequency/acceptability mismatch does not per se question the role of experience for linguistic knowledge, especially if it results from a methodological mismatch. Nor should it be taken to mean that experience cannot be measured in observational data: FREQUENCY OF USE can be operationalized in very distinct ways even for phenomena low in complexity, including – but not limited to – raw, relative, or contingent frequencies. Each of these measures has a different relationship with experimental behavior. This will be illustrated below with *go/come-v* as an example of a construction relatively low in complexity and schematicity: complex measures, not simple frequencies, predict the acceptability of *go/come-v*. Once the methodological mismatch has been reduced, there is a substantial correlation between usage data and acceptability.

It should go without saying that neither corpus analyses nor experiments measure the underlying concepts entrenchment or competence – both are theoretical notions that cannot be accessed directly. Rather, the two methods approach grammatical knowledge from different angles: while experimental tasks tap into the *effects* of entrenchment (or competence), corpus data proxy their *causes* (Stefanowitsch and Flach 2016: 121). The latter assumption is more contentious: corpora are met with substantial skepticism, since they are primarily seen as samples of linguistic output and conventionality (Schmid 2010, 2013). Yet, it is a usage-based assumption that, as samples of production, corpora also represent samples of exposure. This sample is immensely reductive, noisy, and no doubt incomplete. However, the methodological mismatch is a reminder that we need to refine the measurement tools to make appropriate use of corpus data.

The discussion below first describes schema compatibility of *go/come-v* (Section 2) and the construction's distribution in corpora (Section 3). Section 4 reports the acceptability study. Section 5 compares the experimental results with the corpus data, which addresses the question to what extent experimental performance correlates with convention, if an appropriate measure of FREQUENCY OF USE is chosen to approximate the schema compatibility of complex units. The main argument is that while compatibility may be independent of most forms of (raw) frequency, it is highly sensitive to usage-derived properties. Therefore, paying more attention to the multidimensionality of a complex construction can reduce the frequency/acceptability mismatch and account for gradience in acceptability (Section 6). In other words, while acceptability is subject to variation, this variation is congruent with schema compatibility and usage data.

## 2 Schema compatibility

The *go/come-v* construction is suitable for an illustration of the interplay between schemas, usage, and experimental behavior on the one hand, and the frequency/acceptability and methodological mismatches on the other, for two reasons. First, *go/come-v* is subject to a morphological constraint (*Go see the nurse!* vs. *\*He goes sees the nurse*), which has been regarded, at least implicitly, as independent of usage and/or semantics. From a cognitive, usage-based perspective, the construction is primarily subject to a semantic constraint, which affects the likelihood of occurrence in inflectional contexts for functional reasons. This also affects the acceptability of instances that violate only the semantic constraint, but satisfy the morphological constraint. *Go/come-v* is thus illustrative of the effects of schema (in)compatibility. Second, given a low type frequency of two, *go/come-v* is low in schematicity, which makes the construction accessible for the illustration of the methodological mismatch between FREQUENCY OF USE and the measure frequency. This section first discusses how the formal constraint on *go/come-v* can be accounted for in a usage-based model.

The Bare Stem Constraint (BSC) describes the phenomenon that *go/come-v* is grammatical in bare form, but ungrammatical if either verb is (overtly) morphologically marked (e.g., Bjorkman 2016; Jaeggli and Hyams 1993; Pullum 1990). The construction is possible only if *go* or *come* occur as non-third singular indicatives (1a), imperatives (1b), subjunctives (c), or infinitives (1d–e); it is ungrammatical with non-plain forms (third singular, preterite, participles), cf. (2):<sup>1</sup>

---

<sup>1</sup> The participle of *come* in (2e) is identical to the bare form, but still ungrammatical (cf. below).

- |     |    |   |                       |
|-----|----|---|-----------------------|
| (1) | a. | Every day I <i>go get</i> the paper.        | non-3RD.SG indicative |
|     | b. | <i>Go get</i> the paper!                    | imperative            |
|     | c. | She insisted he <i>go get</i> the paper.    | subjunctive           |
|     | d. | I expected him to <i>go get</i> the paper.  | infinitive            |
|     | e. | He doesn't <i>go get</i> the paper.         | infinitive            |
|     |    |   |                       |
| (2) | a. | *She <i>goes gets/get</i> the paper.        | 3RD.SG indicative     |
|     | b. | *They <i>went get/got</i> the paper.        | preterite             |
|     | c. | *We are <i>going getting/get</i> the paper. | ing-participle        |
|     | d. | *They have <i>gone got/get</i> the paper.   | perfect               |
|     | e. | *They have <i>come got/get</i> the paper.   | perfect               |

In formal frameworks, the BSC is accounted for as the result of morphosyntactic parameter or feature operations (Bjorkman 2016; Jaeggli and Hyams 1993). By contrast, a usage-based analysis argues that the morphological constraint follows from a *semantic* constraint (Flach 2015, forthcoming).

What is the semantic constraint that allows *I told her to go get the paper*, but not *\*She goes gets the paper*? It is true that usage-based models have an inherently difficult task to explain ungrammaticality or why something does *not* occur. Usage data only contain positive evidence, so an experience-based view must turn to functional motivations for structural constraints based solely on the commonalities in positive evidence (for a similar idea on island constraints, see Ambridge and Goldberg 2008).

One thing that the vast majority of *go/come-v* have in common is the pragmatic situations in which they occur. As the examples in (3) illustrate, *go/come-v* have a conspicuous preference for orders, suggestions, invitations, or recommendations. These functions are directly encoded in imperatives (3a) or subjunctives (3b), but also occur in the leftward environment: infinitive *go* and *come* are complements of *lets-* or *why-adhortatives* (3c–d), requestive matrix verbs such as *ask*, *force*, or *tell* (3e), or deontic (semi-)modals such as *should*, *must*, *have to*, or *need to* (3f–g):<sup>2</sup>

- |     |    |  |                  |
|-----|----|--|------------------|
| (3) | a. | <i>Go look</i> it up. He brought it up in the primary. [SPOK]                    | imperative       |
|     | b. | ... he called her, <u>insisted that</u> she <i>come eat</i> with us ...<br>[FIC] | subjunctive      |
|     | c. | Thirteen cents to spare. <u>Let's go do</u> some holiday cooking. [SPOK]         | lets-adhortative |

2 Corpus examples are cited from the *Corpus of Contemporary American English* (COCA). Semi-modals are modal expressions that do not satisfy the NICE properties of the core modals (*must*, *should*, *will* etc.), but which are semantically parallel to them (e.g., *have to*, *need to*, *be going to*).



- |    |  |             |
|----|--|-------------|
| d. | Winton replied, “ <u>Why don’t you go build</u> one yourself?” | <i>why-</i> |
|    | [MAG]  | adhortative |
| e. | He <u>asked Monty to go close</u> the back of the truck. [FIC] | requestive  |
| f. | We <u>should go talk</u> to the nurse. [FIC]                   | modal       |
| g. | “I <u>have to go get</u> him,” I told her. [FIC]               | semi-modal  |

Considerably less frequent are uses like in (4), where directive force is absent: (4a) is a general characterization, while (4b) and (4c) describe habitual actions:

- |     |    |  |                      |
|-----|----|--|----------------------|
| (4) | a. | It’s <u>fun to go blow off</u> a little steam afterward ... [SPOK]             | <i>to-complement</i> |
|     | b. | How often do you <u>go see</u> her? [SPOK]                                     | <i>do-support</i>    |
|     | c. | I take them to school, <u>go play</u> golf and pick them up from school. [MAG] | <i>indicative</i>    |

In the terms of classic Speech Act Theory (Searle 1976), orders, suggestions, and recommendations in (3) are non-assertive directives and commissives, while the examples in (4) are representatives and therefore assertive. That is, the events referred to by *go/come-v* are prospected, not asserted (“world-to-word direction of fit”, Davies 1986; Searle 1976). The future-time implication is motivated, although not fully predictable, by the motion verbs *go* and *come*. The directive-commissive function is clear for imperatives and adhortatives. However, for infinitival *go* and *come*, the mapping between syntactic category and pragmatic context is less straightforward: they may occur in directives in (3), but also in representatives in (4). The fine-grained subdivision of infinitival *go* and *come* is motivated by phraseology and goes beyond traditional syntactic categories.<sup>3</sup>

Paying attention to the leftward environment of *go/come-v* recognizes communicative circumstances, which are key to describe the schema and understand the functional motivation of the BSC. The approach is based on the assumption that “you shall know a *construction* by the company it keeps”: the constructional meaning of *go/come-v* can be inferred in much the same way as the meaning of words can be inferred from their collocational behavior (Firth 1957; Harris 1954). As we will see in Section 3, the directive contexts in (3) account for 86.4% of *go-v* and 90.3% of *come-v* in the *Corpus of Contemporary American English* (COCA; Davies 2008). This distribution is more than merely a consequence of the

<sup>3</sup> Note that it is irrelevant whether the verb in the matrix clause is representative: *ask* is assertive in *They asked him to come fix the sink*, but *come fix* is non-assertive relative to the “asking scene”, i.e., the point of the indirect request (cf. below).

BSC – it is a meaningful distributional property that identifies *go/come-v* as a directive, non-assertive construction.

From a Construction Grammar perspective, *go/come-v* qualifies as a form–meaning pair on both dimensions of the original definition (Goldberg 1995: 4). Formally, the BSC is neither predictable from *go* or *come*, which are otherwise not morphologically “defective”, nor from any other construction in English. Semantically, the functional constraint is also not predictable, although it is motivated by the motion verbs that imply futurity.

Recall that schemas arise as abstractions over repeated usage events of similar instantiations. Properties of similarity pertain to form, function, semantics, and pragmatics and involve rich conceptual content. While *schema* (in a Langackerian sense) and *construction* (in a Goldbergian sense) are often used synonymously, they are understood in this context with the following difference: both directive *Go get the paper!* and assertive *I go get the paper* are instances of the same *construction*, but they are not equally sanctioned by the *schema*. As we will see below, an imperative satisfies all functional and pragmatic properties that are associated with the *go/come-v* schema, whereas the indicative does not. In brief, *I go get the paper* is less compatible with the schema.

Compatibility between an instance and the schema depends on how well the instance conforms with the schema’s specifications (Langacker 1987, 1988, 2000). While Langacker makes a distinction between full and partial sanction, sanction is ultimately gradable: partial sanction implies a greater distance from the licensing schema as more specifications are violated (see also Flach forthcoming; Langacker 2000: 12).

An instance is fully sanctioned if it involves a MANDATOR, who requests or suggests, and a MANDATEE, who receives the request or suggestion. The scene is temporally bi-partite: the mandating speech act precedes the mandated state of affairs. In imperatives and adhortatives, all mappings are direct: MANDATOR and MANDATEE map onto speaker and hearer, respectively, the mandating speech act coincides with the moment of speaking and the mandated event is in the future. This also holds in adhortatives, where MANDATOR and MANDATEE are co-referential (*Let’s go have lunch*, *Why don’t we go have lunch?*). Instances are also compatible if configurations change slightly but retain a general alignment with the schema specifications. For instance, while requestive matrix clauses can map MANDATOR and MANDATEE directly onto speaker and hearer, respectively (*I’m telling you to go see the nurse*), situations can shift along the temporal axis and/or extend to third parties (*I/They told him to go see the nurse*). In passives, the encoding of the MANDATOR is absent (*She was told to go see the nurse*). Usage events with imperatives (IMP), adhortatives (WHY, LETS), and requestive patterns (REQ) are fully

compatible with the schema, because all participants and mappings are available or pragmatically understood.

Note that two instantiations of the same context may differ in compatibility. For instance, the adhortative *why don't **we** go see a movie?* is a suggestion between speaker and hearer, whereas *why don't **they** go see a movie?* can also be interpreted as an enquiry about a potential or missed option by a third party. Similarly, the compatibility of (semi-)modals depends on the type of modality: deontic expressions (MOD: *must*; SEMI: *have to*) are directives, although the deontic source is not mapped.

Sanction is only partial in situations which lack a MANDATOR–MANDATEE arrangement. Future-time expressions (*will*, *going to*) are not directive, but they are non-assertive by encoding intention, imminence, or prediction. By contrast, *to*-complements (TO.COMP: *it's fun to go blow off steam*), indicatives (IND: *I go see my lecturer often*), and *do*-support (DO: *Do you go exercise regularly?*) lack all relevant configurations. That said, the leftward nominal patterns in *to*-complements often invite future readings of the content clause (*It's a chance to come see history*). The vast majority of actually occurring indicatives are strongly biased towards futurate uses (*we go clash here tonight*) or non-assertive conditionals (*If you go see him again,...*). That is, truly assertive *go/come-v* of the type *I go get the paper every morning* are rare in corpora.

In sum, the directive environments (IMP, WHY, LETS, and REQ) satisfy a semantic constraint: they map all scene and participant configurations of the higher-order schema (Langacker 1988: 132). By contrast, the contexts IND, TO.COMP, and DO violate the semantic constraint: they are extensions with increasing distance from the schema (MOD and SEMI are somewhere in the middle). While extensions may “pass unnoticed in normal language use” (Langacker 2000:17), their compatibility is functionally compromised, which predicts lower likelihood of occurrence and lower acceptability.

As with any gradable concept, it should be borne in mind that individual cases may escape a clear classification as fully or partially sanctioned. But the categorization by syntactic environments captures the main idea: the coding of only the syntactic environment covers the corpus data remarkably well (cf. Section 3).

Compatibility relates to similarity in the following way: a commissive *let's go see a movie* is similar to another commissive *why don't we go see a movie*, but not by a direct comparison of form, but via a semantic-functional overlap. Although *Go see the doctor!* is structurally dissimilar to *I told her to go see the doctor*, they are similar because the participant configurations of directives are present in both. In other words, abstracted schemas maximize relevant similarities, but minimize the effects of idiosyncratic details. Most importantly, schemas may go well beyond

structural or formal properties and involve rich, abstract extra-grammatical knowledge.

Let us return to the question from the beginning of this section: how can we use positive evidence to account for why inflectional *go/come-v* does not occur? The key is that the inflectional contexts in English (preterite, progressive, and perfect) encode representatives: *\*She went saw the doctor*, *\*We were going eating there last night*, or *\*They have come water the plants* are not directive and cannot encode a MANDATOR–MANDATEE configuration. They are highly *improbable* to occur with *go/come-v* for functional reasons. In other words, the morphological constraint follows the semantic constraint because of the (non-causal) correlation between the way directives are expressed in English and the English morphological paradigm.<sup>4</sup> This also explains the ungrammaticality of *\*They have come water the plants* despite the (surface) bareness of *come*: it can be linked directly to the inability of English perfects to express non-assertive content. This explanation of the BSC does not require an elaborate integration of the accidental form syncretism of *come* (which in all likelihood has nothing to do with *go/come-v*). Put simply, whether *come.PRT* is “featurally inflected” is irrelevant for the usage-based explanation of the BSC.

### 3 Corpus distribution

From the discussion in the previous section, we can expect that schema compatibility is reflected in corpus data in two ways. First, fully compatible contexts (IMP, WHY, LETS, REQ) should occur with *go/come-v* more frequently than expected, while the reverse should hold for semantic constraint violators (SEMI, IND, DO, TO.COMP). Second, there should be a continuum from imperatives (as the most compatible) to indicatives (as the least compatible), reflecting an increasing distance from the schema. We will conclude the section with a discussion of usage data with respect to the tension between “potential space” and “instantiated space” (cf. Langacker 2000).

#### 3.1 Corpus data

Table 1 gives the distribution of 1000 random observations for each construction type (CXN) of *go/come-v* and their coordinated alternations across syntactic environments.<sup>5</sup> The data were extracted from the 2015 offline version of the *Corpus of*

---

<sup>4</sup> Where they do occur, they are modeled on bare indicatives (e.g., *went-v*; Flach 2015).

<sup>5</sup> The semantically related *go/come-and-v* are not subject to the BSC, but they are also skewed towards bare forms (Flach forthcoming); the coordinated types lie on a continuum of *go/come-(and)-v* constructions (cf. Figure 2).

*Contemporary American English* (COCA; for query details, see Flach forthcoming). All data points were coded for their syntactic environment (SYN), that is, either the syntactic category for imperative or indicative *go* and *come* (IMP, IND) or the leftward context for infinitival *go* and *come* (WHY, LETS, REQ, MOD, TO.COMP, DO). As subjunctives (*They recommended he go see a doctor*) are too rare, they are subsumed under REQ. See the Appendix for a summary of query information, categories and examples.

The table also contains a set of 1000 random tokens of bare verbs which could have occurred with *go/come-v* (i.e., not an auxiliary or an instance of one of the four other patterns). This represents the average corpus use and acts as a control sample. We use the sample to determine by how much the co-occurrence of *go/come-v* with a syntactic environment deviates from expectation under the assumption that there is no relationship between constructions and environments. Recall that directive contexts should occur with *go/come-v* more often *than expected*. This is not necessarily obvious from simple counts. For instance, what does it mean that 259 of 1000 of *go-v* uses are imperatives? If a quarter of bare verb uses in COCA were imperatives, this rate would not be noteworthy, because *go-v* would occur in the imperative *as often as expected*. However, since the average rate of imperatives in the control sample is 6.1%, the imperative rates for *go-v* (25.9%) and *come-v* (37.4%) deviate substantially from this expectation. Conversely, the indicative rates for *go-v* (5.9%) and *come-v* (3.1%) are well below the control level (27.1%).

The table shows the preference of *go/come-v* for requestive environments (IMP, WHY, LETS, REQ, MOD, and SEMI), which account for 86.5% of *go-v* and 90.3% of *come-v*, but only for 47.4% of the control sample. Note that while *go/come-v* are

**Table 1:** Distribution of *go/come-(and)-v* across syntactic environments (1000 tokens per CXN).

SYN	<i>go-v</i>	<i>come-v</i>	<i>go-and-v</i>	<i>come-and-v</i>	control	Dim 1 CA <sub>SRC</sub>
IMP	259	374	108	186	61	1.255
WHY	14	11	9	9	1	1.114
LETS	80	0	24	0	9	0.775
REQ	133	205	143	276	102	0.465
SEMI	173	97	194	93	68	0.302
MOD	206	216	268	252	233	-0.094
TO.COMP	67	59	116	71	180	-1.104
IND	59	31	110	97	271	-1.778
DO	9	7	28	16	75	-2.230
TOTAL	1000	1000	1000	1000	1000	

restricted to deontic modals, the MOD category in the control sample does not distinguish epistemic and deontic modals. This may underestimate the relevance of MOD for *go/come-v* in this distribution (cf. the relatively equal MOD values across the table).

The order of rows in Table 1 represents an increasing distance from the schema as discussed in Section 2. This order is confirmed by a Correspondence Analysis (CA), as shown by the Dim 1 CA<sub>SRC</sub> values in the last column. CA helps detect trends in complex tabular data that are difficult to eye-ball; we'll discuss the method in the next section.

### 3.2 Correspondence analysis

Correspondence Analysis (CA) is a dimension reduction method for categorical variables that aims to detect patterns in multidimensional tabular data like in Table 1.<sup>6</sup> While the mathematical background of CA is beyond the scope of this paper, the conceptual idea is relatively simple (see Greenacre 2017 for an accessible introduction; for corpus applications, see Glynn 2014; Levshina 2015). In a nutshell, the Dim1 CA<sub>SRC</sub> values in Table 1 capture the (dis)similarity of a row relative to all other rows. For instance, the row profile, or vector, of IMP = [259, 374, 108, 186, 61] is more similar to WHY = [14, 11, 9, 9, 1] than to IND = [59, 31, 110, 97, 271], despite the differences in frequency between IMP and WHY. This is intuitive since the higher values cluster at the beginning of the IMP and WHY vectors, but at the end of the IND vector. In technical terms, the vectors are the syntactic environments' coordinates in a five-dimensional space.

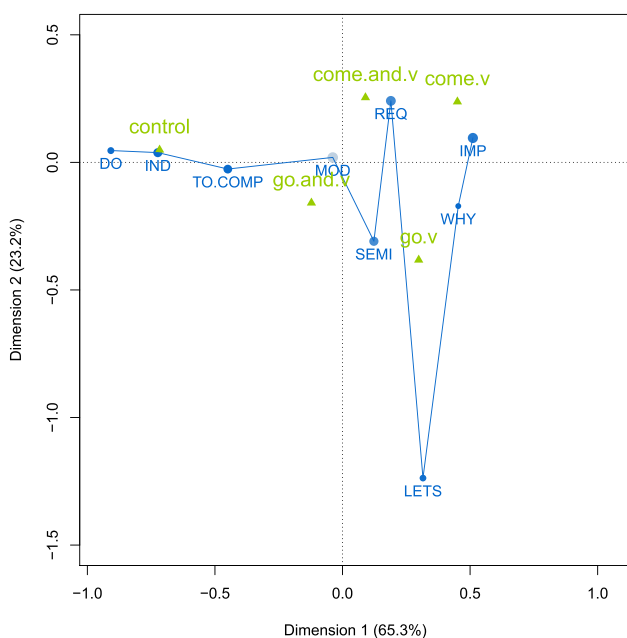
Since a five-dimensional space is difficult to imagine, let alone visualize, CA reduces the complexity in a matrix (like Table 1) to a few interpretable dimensions. After reduction, the first dimension explains most of the variance. For the rows, this is represented by the standard row coordinates in Table 1 (Dim 1 CA<sub>SRC</sub>). Given their distribution across the network, the CA<sub>SRC</sub> values best distinguish between the syntactic environments without losing too much information. Put simply, they quantify the rows' (dis)similarities in terms of the columns, that is, they

---

<sup>6</sup> The sampling of 1000 observations per CXN does not reflect that *go-v* (13,049) is almost four times as frequent as *come-v* (3528). However, CA works on proportions rather than raw frequencies. Sampling is preferable because devising a numerically matching control sample is impossible. However, the fact that the CA values of several CAs with all 24,533 *go/come-(and)-v* data points and differently sized control samples were highly correlated, suggests that the current sample is sufficiently robust.

approximate the environments' association with *go/come-v*.<sup>7</sup> (The same logic holds for the difference between *column* profiles, i.e., (dis)similarity between constructions.)

$CA_{SRC}$  values can be conceptualized as a continuum. This becomes more tangible if we discuss them together with the plot in Figure 2. So-called biplots are the visual representation of a high-dimensional data structure (Table 1) in a two-dimensional space (Figure 2). Since the biplot captures 88.6% of the variance (65.3% for the first dimension on the *x*-axis and 23.3% for the second dimension on the *y*-axis), we can interpret the 2D representation with reasonable confidence.



**Figure 2:** CA biplot of the *go/come(-and)-v* network and the control sample. Dot size represents frequency and color depth represents distinctiveness (e.g., MOD is frequent, but not distinctive; WHY is distinctive, but not frequent).

<sup>7</sup> From a data set like in Table 1, the computation of  $CA_{SRC}$  involves the calculation of observed and expected proportions as well as indexed and standardized residuals that measure the deviance of the observation from the expectation (not unlike in a  $\chi^2$ -test). The matrix of standardized residuals is reduced by Singular Value Decomposition (Greenacre 2017: Appendix B). In praxis,  $CA_{SRC}$  are returned by the `ca()` function in the package `{ca}` (Nenadić and Greenacre 2007). See <https://osf.io/a293t/> for step-by-step R code to compute the  $CA_{SRC}$  “by hand”.

The bi-plot simultaneously shows row profiles (SYN, blue) and column profiles (CXN, green): distributionally related categories populate similar plot regions. The closer a category member is to the center at  $[0,0]$ , the less distinctive it is for the data overall (e.g., MOD) and vice versa. For instance, LETS is furthest from the center because its row profile deviates most markedly from all other row profiles. This can be inferred intuitively from Table 1, where LETS is near-exclusive to *go-(and)-v*. Hence, its nearest neighbors are the *go*-patterns. Note, however, that numerical distances between row and column labels are not meaningful in CA. This may be counterintuitive, but the fact that LETS is very far from *go-v* does *not* mean that *go-v* is more strongly associated with WHY or SEMI. Rather, the “outlier” position of LETS reflects the fact that LETS is *only* associated with *go-(and)-v*, so it is furthest from all other constructions.

Conceptually, we can think of the plot as a map or a “constructional ecology” (Taylor 2004): *go/come-v* occur in non-assertive environments in the right quadrants (IMP, LETS, WHY, REQ, WHY). Average corpus use, on the other hand, occurs relatively more often in assertive environments in the left quadrants (IND, DO, TO.COMP). Overall, the plot shows two important continua from left to right along Dimension 1. First, the constructional continuum runs from average corpus use via the coordinated types to the serial types. The *come*-types are more “extreme”, i.e., always further right, than the corresponding *go*-types. Second, the directive continuum, which is more important for our current purpose, runs from assertive (DO, IND) to directive environments (LETS, IMP), indicated by the blue line.

Let us return to the  $CA_{SRC}$  values in Table 1. They correspond to the syntactic environments’ values on the x-axis of the plot.<sup>8</sup> In other words, even though the relationship between the rows in Table 1 is high-dimensional, it can be represented as a one-dimensional continuum (Figure 3). This representation is without a significant loss of relevant information, if the information we are interested in is the assertive–directive continuum. This continuum underlies the order of the rows in Table 1 and corroborates the discussion in Section 2.



**Figure 3:** One-dimensional representation of the assertive–directive continuum (Dim1  $CA_{SRC}$ ).

<sup>8</sup> We say “correspond” because the  $CA_{SRC}$  value in Table 1 are the dimension 1 *standard* row coordinates ( $CA_{SRC}$ ), while x-axis values in Figure 1 are the dimension 1 *principal* row coordinates ( $CA_{PRC}$ ).  $CA_{SRC}$  and  $CA_{PRC}$  are perfectly correlated ( $r = 1$ ). This study uses  $CA_{SRC}$  for practical reasons because they are returned as the output of the `ca()` function in the R {ca} package (Nenadić and Greenacre 2007), while  $CA_{PRC}$  are calculated only during plotting (cf. <https://osf.io/a293t/>).



Before we move to the judgment experiment, we briefly return to and discuss the increasingly complex ways in which FREQUENCY OF USE can be measured for *go/come-v* (cf. Section 5), although all measures are based on the same data in Table 1. The simplest is the corpus frequency ( $F_{\text{CRP}}$ ) of the environments, which we may express as a percentage of their occurrence in the control sample (e.g., 6.1% IMP, 27.1% IND). A more complex measure is the frequency of a syntactic environment in the construction ( $F_{\text{CXN}}$ ), expressed as the average of co-occurrence with *go/come-v* (e.g., 31.7% IMP, 4.5% IND). This is more complex than corpus frequency, because the  $F_{\text{CXN}}$  value of an environment depends on the  $F_{\text{CXN}}$  value of all other environments in the same constructional space.

The most complex measure is  $\text{CA}_{\text{SRC}}$ , not necessarily because of its computation, but because it reflects, simultaneously, the association between a syntactic environment and *go/come-v* relative to all other syntactic environments *and* constructional alternatives including the control sample.  $\text{CA}_{\text{SRC}}$  does not derive straightforwardly from – or correlate with – either  $F_{\text{CRP}}$  or  $F_{\text{CXN}}$ : the two most strongly associated directive environments LETS and WHY are amongst the most infrequent, both in the corpus and in the construction. However,  $\text{CA}_{\text{SRC}}$  is frequency-related in so far as it is based on co-occurrence *across* a table, that is, it is contingent on the frequency of contexts in other parts of the network (and the corpus in this case). This is the underlying logic of all association measures. Frequency (counting) and association (distribution) can be two very different things, although they often correlate. Note that the increasing complexity in terms of association goes far beyond the question of choosing the right granularity (“abstractness”) of the syntactic unit being counted in a corpus (Bader and Häussler 2010; Crocker and Keller 2006).

### 3.3 Discussion

The distributional analysis of the constructional network as an ecological space reflects the idea that a construction’s *potential space* and its *instantiated space* are not the same thing (Langacker 2000: 29). Conventional use, i.e., instantiated space, will cluster in particular regions of the potential space (Langacker 1988: 153, 2000: 31). One of the main differences between formal and usage-based models is that the former are more (or solely) interested in potential space, while the latter place more emphasis on instantiated space. Inferring negative evidence via association contrasts potential with instantiation: the assumption is that speakers are sensitive to the quantitative distributional tension between potential and instantiated space, so that acceptability (and arguably grammaticality) diminishes the further a usage event is from the conventionally instantiated space.

The usage-based analysis of *go/come-v* and its distribution in corpora show two things. First, the relationship between morphology, semantics, and pragmatics is intimately related such that a binary distinction into grammatical and ungrammatical *go/come-v* may be too simple. If distributional information is added to the picture, the morphological constraint can be shown to follow from the semantic constraint. The semantic constraint can be violated more easily, as long as the expansion of the instantiated space has communicative value and does not violate an entrenched morphological constraint. (One could say that *I go get the paper every morning* and *Did he go eat there?* “piggy-back” on the entrenchment of bare *go/come-v*.) By contrast, since inflected contexts have no potential for the expression of non-assertiveness, violating the morphological constraint has no functional motivation. This non-occurrence contributes to the negative entrenchment of inflectional forms and continues to constrain the potential space.<sup>9</sup> The distinction between a floutable semantic constraint and an absolute morphological constraint is similar to the distinction between soft and hard constraints in formal syntax (Keller 2000; Sorace and Keller 2005).

Second, although corpus data only hold what does occur, positive evidence can yield insights into ungrammaticality by statistical inference (Stefanowitsch 2008). The inference is based on contingency, because we have no way of knowing from raw frequency alone whether a preference deviates from expectation (or what the expectation is in the first place). Furthermore, although the operationalization of schema compatibility of *go/come-v* is more complex than raw frequency, it is empirically relatively simple, since the pragmatic situation is only represented by the syntactic environment. That said, the annotation scheme is based on the investigation of vast amounts of usage. In addition, it required the refinement of a traditional syntactic category (infinitive) on phraseological grounds to capture constructional semantics, which some may find unconventional or even unwarranted.

As for the acceptability experiment, there is little to go by from the previous literature. Given their focus on (un)grammaticality, the expert judgments are binary and expectedly unanimous (Bjorkman 2016; Jaeggli and Hyams 1993; Pullum 1990). On the other hand, an informal acceptability survey with naïve participants reports that a fifth of respondents rejected bare indicatives on a binary choice

---

<sup>9</sup> This is not to say that this situation is fixed. A significant rise or drop in one section of the instantiated space, e.g., in imperative or adhortative uses, can shift the schema towards lesser or greater assertiveness (which is what happened to *go/come-v* in Middle and Late Modern English, respectively, cf. Flach forthcoming), potentially making it more or less amenable to inflectional contexts. A greater degree of assertiveness explains why the coordinated *go/come-and-v* are not subject to the BSC (Flach forthcoming).

(Pullum 1990), which suggests that the semantic constraint does affect the acceptability of *go/come-v*.

From the discussion above, we expect that acceptability follows schema compatibility on two levels. First, contexts that satisfy the semantic constraint (IMP, LETS, WHY, REQ) will be judged more acceptable than those which violate it (SEMI, TO.COMP, IND). Second, the acceptability of the semantic constraint violators is predicted to be gradual, corresponding to their distance from the schema and following usage distribution (i.e., SEMI > TO.COMP > IND). Simple pattern frequency ( $F_{CRP}$ ) would predict roughly the opposite (IND > ... > IMP/WHY/LETS). The prediction from construction frequency ( $F_{CXN}$ ) is mixed, but we would expect higher ratings for frequent contexts IMP, REQ, or TO.COMP than for rare LETS or WHY.

## 4 Experiment

### 4.1 Materials

Thirty sentence pairs were created, i.e., 15 per construction type (CXN: *go/come*), of which 12 pairs were in bare form (IMP, LETS<sub>GO</sub>/WHY<sub>COME</sub>, REQ, SEMI, TO.COMP, IND). LETS was used with *go* and WHY with *come*, because *come* cannot felicitously occur with LETS (<sup>2</sup>*Let's come play tennis*; cf. Table 1). Three pairs with inflections (PST, PRT, THIRDPERS) ensured that participants used the full range of the rating scale.

The lexical material in the bare items was controlled for *go/come-v* association: one member of each pair contained an associated  $V_2$  and the other a non-associated  $V_2$ . Association was determined by two Simple Collexeme Analyses (SCA; Stefanowitsch and Gries 2003) over the COCA data of 13,050 *go-v* and 3528 *come-v* observations using the R package {collostructions} (Flach 2017a). Since SCA returns statistical significance quickly for low-frequency verbs in large data sets in large corpora, the absence of an association was defined as collostruction strength below  $G^2 \leq 6.64$  ( $p \geq 0.01$ , \*\*). Highly associated verbs that are either part of an idiom (*go figure*, *go fish*) or potentially insulting (*fuck*, *kill*, *pee*, *hang*, etc.) were excluded. The pairs were structurally identical, with only a minor lexical adjustment in one pair to avoid the violation of selectional restrictions (*watch movies* vs. *hear stories*).

- (5) a. **Go** { *find* | *seek* } help immediately! { associated | non-associated }  
 b. Dad invited them to **come** { *stay* | *live* } with us. { associated |  
 non-associated }

For the bi-clausal contexts REQ and SEMI, characteristic left-context types were selected based on their frequency with *go/come-v* in COCA (*tell/ask NP to go-v; invite/call NP to come-v; have/need/going to go/come-v*). For TO.COMP and IND, one item each implied futurity and the other habituality or stativity:

- (6) a. It's a chance to come { **help** | **support** } a friend. motion  
 b. Children like to come { **hear** stories | **watch** movies }. no motion  
 (habitual)
- (7) a. I go { **fetch** | **retrieve** } the mail regularly. motion  
 b. They go { **sleep** | **relax** } on the couch. no motion (stative)

The inflectional pairs only contained associated  $V_2$  and were varied by inflection on  $V_2$  (*went bought/buy, goes speaks/speak*).<sup>10</sup> To ensure that subjects were aware of the formal range, five training items covered the full morphological paradigm.

The sentence pairs were split into two lists (30 sentences, 15 per CXN; 18 with associated and 12 with non-associated  $V_2$ ). Each  $V_2$  occurred only once per list and subject NPs were balanced between pronoun and full NPs.<sup>11</sup> The lists were pseudo-randomized and set up in two orders, one in reverse of the other, to produce a total of four questionnaire versions.

The sentences in each version were set up five to a page in Qualtrics.<sup>12</sup> Each page contained a maximum of three *go* or *come* sentences and a maximum of three associated  $V_2$ . Each page contained one inflected context and at least one fully compatible context (IMP, LETS, WHY, or REQ), but no context occurred twice in a row and no more than once per page. The sentences were centered above a horizontal scale with numerical values from 1 to 7. The endpoints had categorical labels (“unacceptable” and “perfect”, respectively). All pages had a forward button only.

Participants were instructed to rate the naturalness of the sentences in informal conversations with family, friends, or co-workers. On the final page, a text field required them to state their suspicion on the purpose of the study, but they were allowed to enter a single character if they had none. Two optional text fields asked them to type age and gender.

<sup>10</sup> As inflectional contexts were not of interest for the current question and used primarily as anchors for subjects to use the full range of acceptability, they were not systematically controlled for  $V_2$ .

<sup>11</sup> *You* is more common in *come-v* items than other pronouns to avoid unrelated oddness due to *come*'s path semantics towards the deictic center (?*I/we come fix the sink*). Conversely, *you* is not used with indicatives to avoid ambiguity with imperatives (i.e., it would be unclear whether participants judge the imperative or indicative reading of *You go get the mail*).

<sup>12</sup> Qualtrics, Provo, Utah/USA, version 02/2018, available at [www.qualtrics.com](http://www.qualtrics.com).

## 4.2 Participants

Forty participants were recruited through Prolific Academic (Palan and Schitter 2018). The platform's pre-screening options ensured that the survey was only available to English monolingual L1 speakers aged 18–50, who were born in the US or Canada, currently reside in their country of birth, and have spent a maximum of 6 months in a foreign country. All participants were required to have a Prolific Academic approval rate of 100%, which quantifies their cooperative behavior in previous studies (cf. Häussler and Juzek 2016).

The participants (15 female, 22 male, 2 na; mean age 30.8,  $sd = 8.7$ ) were awarded £0.80/\$1.26 for the evaluation of 35 sentences, which took an average of 4 min. Participants were randomly assigned to one of the four versions, but all started with the same page of training items, which were shown in random order. A total of 1198 ratings were collected (two participants did not provide ratings for one item each).

## 4.3 Analysis

Most studies in Experimental Syntax treat Likert-type responses as numerical and fit linear regression models (Sprouse et al. 2013; Weskott and Fanselow 2011), while ordinal regression remains true to the nominal character of the responses (Baayen and Divjak 2017; Endresen and Janda 2017). We discuss the results of a Linear Mixed-Effects Model for numerical data (LMER; Baayen et al. 2008), which was more sensitive to interactions than a Cumulative Link Mixed Model for ordinal data (CLMM; see Endresen and Janda 2017). However, both models produce essentially identical results, as their coefficients are near-perfectly correlated ( $r = 0.99$ ,  $t = 21.7$ ,  $p < 0.0001$ ).

Since an LMER treats responses as numerical, the ratings can be *z*-scaled to normalize scale compression. This balances out differences in subjects' interpretation of the Likert scale, because not all participants exhaust the full 1–7 range: some restricted their judgments to the 5–7 region, others to 2–6, etc. (Coward 1997; Schütze and Sprouse 2013; Sprouse et al. 2013). All ratings were *z*-scaled using the formula in Schütze and Sprouse (2013: 43), so that “each response [by a participant *P*, *SF*] is expressed in standard deviation units from *P*'s mean”. This transformation retains the differences *within* a participant's responses, but makes responses comparable *across* participants.

The variables syntactic environment (SYN; levels: IMP, WHY, LETS, REQ, SEMI, TO.COMP, IND) and  $V_2$  association (ASSOC; levels: YES, NO) were included as fixed effects, and SUBJECT, VERB, and LENGTH as random effects. LENGTH (in words) is

included because IMP, LETS, and IND are shorter than bi-clausal REQ or TO.COMP; but the length of a sentence is difficult to control if we want to avoid artificially complex and potentially unidiomatic adjuncts. ITEM is the manipulated variable of interest and is therefore not included as a random effect (cf. Sprouse et al. 2013: 226). In a full model, CXN type (*go, come*), AGE and GENDER had no effect and were removed.

The frequency-related corpus measures cannot be included in the regression, because they are not free to vary with SYN: that is, all data points of a given level of SYN have identical values for CA,  $F_{\text{CXN}}$ , and  $F_{\text{CRP}}$ , so these measures do not discriminate. The relationship between acceptability ratings and corpus measures will be discussed in Section 5.

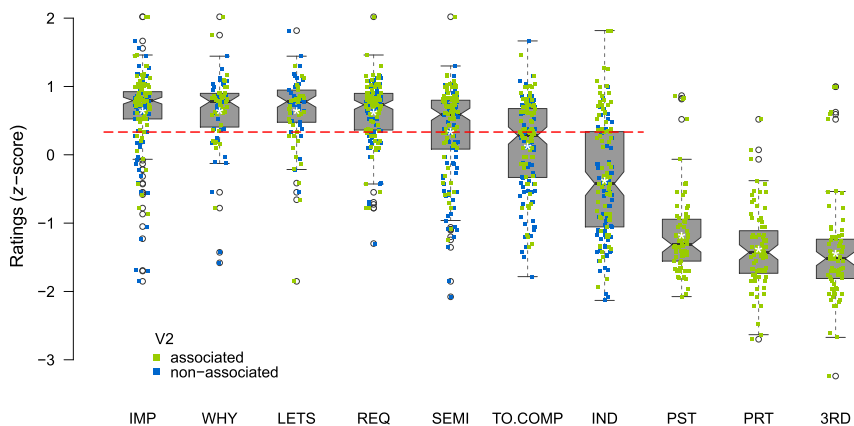
4.4 Results

The ratings are summarized in Table 2 and plotted in Figure 4. The orderings from top to bottom (table) and left to right (figure) reflect increasing schema distance. The compatible contexts IMP, WHY, LETS, and REQ have the highest medians and means; they also tend to be lower in variance. Expectedly, the inflectional contexts receive the lowest ratings. There is no significant difference between *go* and *come* ( $go\text{-}v = -0.029$ ;  $come\text{-}v = 0.030$ ; Wilcoxon  $p = 0.39$ ).<sup>13</sup>

**Table 2:** Summary of acceptability ratings of *go/come-v*: Median (over raw ratings), Mean (over z-scores), and Sd (over z-scores).

SYN	Example	Median	Mean	SD
IMP	<i>Go get me some water, please.</i>	7	0.64	0.61
WHY	<i>Why don't you come see me after class?</i>	7	0.64	0.55
LETS	<i>Let's go cook dinner together!</i>	7	0.65	0.52
REQ	<i>Mum asked us to go help our brother.</i>	7	0.62	0.45
SEMI	<i>I'm gonna go send them an email.</i>	6	0.34	0.67
TO.COMP	<i>She likes to go play chess.</i>	6	0.12	0.68
IND	<i>They come fix computers for a living.</i>	5	-0.38	0.84
PST	<i>My sister went bought the groceries.</i>	2	-1.18	0.60
PRT	<i>We had gone watched a movie.</i>	2	-1.38	0.58
3RD	<i>Helen comes speak to her parents daily.</i>	2	-1.44	0.73

<sup>13</sup> For a detailed overview of mean ratings by verb association, see Table 6 in the Appendix.



**Figure 4:** Ratings (z-score) for *go/come-v* by syntactic environment. Scatter dots represent associated (green) and non-associated  $V_2$  (blue); the red line shows mean ratings for all bare environments.

Three clusters emerge. The first comprises the compatible contexts IMP, WHY, LETS, and REQ, which are not judged notably different from each other. In the second cluster with the semantic constraint violators SEMI, TO.COMP, and IND ratings begin to drop and variation rises, especially for IND. The third cluster contains the inflectional contexts, reflecting the hard morphological constraint.

Table 3 shows the results of the LMER model with the interaction between the syntactic environment (SYN) and  $V_2$  association (ASSOC).

The values in column 2 estimate the acceptability of a syntactic environment relative to the baseline IMP (the intercept at 0.797). As the estimates are negative, all other levels of SYN are, on average, judged less acceptable than imperatives. For example, a WHY item is judged lower by 0.364 z-score units than IMP. The differences in ratings are not significant for the compatible contexts IMP, WHY, LETS, and REQ ( $p > 0.05$ ). However, the constraint violators SEMI, TO.COMP, and IND receive significantly lower ratings ( $p < 0.05$ ).

In order to compare the experimental results with corpus measures in Section 5, LMER coefficients were extracted for each level of SYN by subtracting its estimate in column 2 from the IMP baseline of 0.797 (i.e., IMP: 0.797; WHY: 0.433; LETS: 0.776; REQ: 0.505; SEMI: 0.176; TO.COMP:  $-0.056$ ; IND:  $-0.648$ ).

Overall, items with an associated  $V_2$  receive higher ratings ( $mn = 0.43$ ,  $sd = 0.67$ ) than with a non-associated  $V_2$  ( $mn = 0.23$ ,  $sd = 0.79$ ), which is a significant difference (Wilcoxon:  $p < 0.001$ ). However, as the interaction term (SYN\*ASSOC) shows, an associated  $V_2$  influences the acceptability *only* for the constraint violators SEMI, TO.COMP, and IND. For example, a semi-modal with an associated  $V_2$  is judged 0.522 z-score units better than a semi-modal with a

**Table 3:** Summary of the LMER model of acceptability of bare *go/come-v*. Significant variable levels and interactions shown in grey.

	Estimate	SE	t-value	Pr(> t )
(Intercept)	0.797	0.163	4.902	0.0002 ***
Syn = WHY	−0.364	0.273	−1.333	0.19 ns
Syn = LETS	−0.022	0.267	−0.081	0.94 ns
Syn = REQ	−0.292	0.231	−1.266	0.22 ns
Syn = SEMI	−0.621	0.184	−3.382	0.002 **
Syn = TO.COMP	−0.854	0.171	−4.989	8.77e−06 ***
Syn = IND	−1.445	0.199	−7.246	1.76e−08 ***
Syn = WHY:Assoc = yes	0.204	0.343	0.594	0.56 ns
Syn = LETS:Assoc = yes	−0.184	0.341	−0.541	0.59 ns
Syn = REQ:Assoc = yes	0.076	0.258	0.297	0.77 ns
Syn = SEMI:Assoc = yes	0.522	0.214	2.440	0.026 *
Syn = TO.COMP:Assoc = yes	0.435	0.202	2.154	0.032 *
Syn = IND:Assoc = yes	0.670	0.256	2.617	0.012 *

**Random effects:**

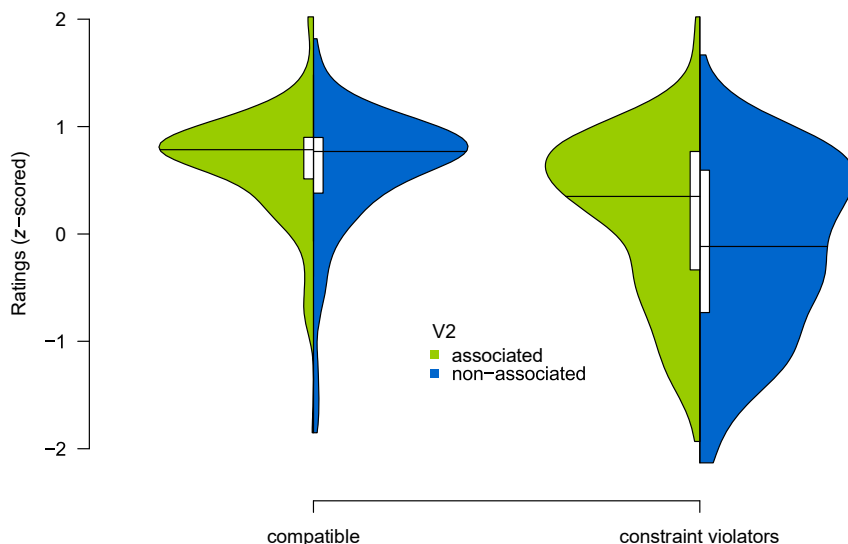
Groups	Name	Variance	Std. Dev.
SUBJECT	(Intercept)	0	0
VERB	(Intercept)	0.076	0.276
LENGTH	(Intercept)	0.036	0.190
Residual		0.340	0.583

Number of obs: 959, groups: SUBJECT 40; VERB 39, LENGTH 5  
Model: lmer(Z.score ~ SYN + SYN\*ASSOC + (1 | SUBJECT) + (1 | VERB) + (1|LENGTH)).

non-associated  $V_2$ . There is no influence of verb association for fully compatible items (IMP, WHY, LETS, REQ). This effect is graphically shown in Figure 5, which aggregates the ratings for compatible and constraint violator contexts.

Finally, the influence of increased schema distance is consistent for the context-specific manipulation of the semantic constraint violators. Recall that half of the items for SEMI are deontic (SEMI<sub>1</sub>: *have/need to*) and the other half are intention-based (SEMI<sub>2</sub>: *going to*). Similarly, one half of TO.COMP and IND items describe situations with implied futurity (TO.COMP<sub>1</sub>: *It's a chance to come help a friend*) or motion (IND<sub>1</sub>: *I go fetch the mail regularly*), while the other express habituals (TO.COMP<sub>2</sub>: *They like to go watch movies*) or statives (IND<sub>2</sub>: *I go sleep on the couch*). For semi-modals, the difference between items closer to the schema (index 1) and those further away (index 2) is not significant (SEMI<sub>1</sub> = 0.37 vs. SEMI<sub>2</sub> = 0.31;  $p = 0.17$ ), but schema distance significantly affects infinitival complements (TO.COMP<sub>1</sub> = 0.33 vs. TO.COMP<sub>2</sub> = −0.09;  $p < 0.001$ ) and indicatives (IND<sub>1</sub> = −0.23 vs. IND<sub>2</sub> = −0.53;  $p < 0.04$ ).





**Figure 5:** The interaction of compatibility and verb association in acceptability judgments.

In summary, the absolute morphological constraint is expectedly confirmed, although there is no sharp drop that might be suggested by a view on *go/come-v* solely in terms of binary (un)grammaticality. The results are broadly compatible with the informal survey where imperatives and infinitives were acceptable, while indicatives were rejected by 20% of naïve respondents on a binary choice (Pullum 1990). Crucially, acceptability depends on whether the semantic constraint is satisfied or not. While so-called soft constraints are known to be sensitive to context manipulation (Sorace and Keller 2005: 1509), the sensitivity for contextual manipulation corresponds directly, and very systematically, to schema distance. In Langacker’s terms, the extensions to semantically incompatible contexts become noticeable “when a conflict is egregious, or when small conflicts have a cumulative effect” (Langacker 2000: 17). The cumulative effect arises with diminishing directive force in the vicinity of *go/come-v*. This pattern is consistent with usage data on the right level of complexity, to which we now turn.

## 5 Comparing corpus and experimental data

The results from the judgment task confirm that acceptability is congruent with schema compatibility, depending not only on the violation of a (hard) morphological constraint, but also on the violation of a (soft) semantic constraint. This

section addresses the relationship between experimental behavior and corpus distribution.

The boxplot in Figure 4 shows that acceptability is at ceiling for four contexts and drops for constraint violators. However, the ordering of the boxplots implies a non-linear relationship between equidistant contexts. The discussion of corpus distribution above showed that this may not be the case (cf. Figures 2 and 3). Hence, Figure 6 plots corpus distribution of the contexts “to scale” on the x-axis ( $CA_{SRC}$ ), which reflects the environments’ distances quantitatively. The relationship between corpus data and acceptability judgments (LMER coefficients) is highly correlated: the closer to the schema, the higher the acceptability ( $r = 0.93$ ,  $t = 5.74$ ,  $df = 5$ ,  $p < 0.01$ ).

We now return to the other usage measures to address the relationship between the frequency/acceptability and methodological mismatches. Recall that FREQUENCY OF USE of *go/come-v* can be measured in three increasingly complex ways: (i) the frequency of the contexts in the corpus ( $F_{CRP}$ ), (ii) their frequency in the construction ( $F_{CXN}$ ), and (iii) their association with the construction ( $CA_{SRC}$ ). Table 4 shows the correlation of these measures with the coefficients from the LMER model.

First, acceptability is strongly correlated with  $CA_{SRC}$ . Second, raw corpus frequency ( $F_{CRP}$ ) in the final column shows the exact opposite. In other words, the high frequency of IND in a corpus (27.1%) does not “save” the context from low acceptability with *go/come-v*. Conversely, the lower frequency of WHY or LETS in the corpus and the construction does not entail low acceptability of *go/come-v*. Third,

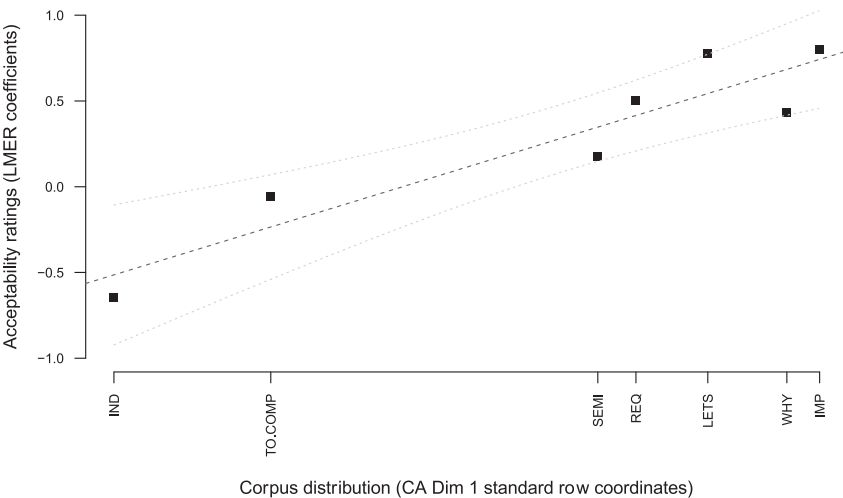


Figure 6: Corpus distribution vs. acceptability ratings (lm:  $R^2 = 0.87$ ; with 95% CI).

**Table 4:** Correlation of LMER coefficients with usage-derived measures (COCA).

SYN	LMER	CA <sub>SRC</sub>	F <sub>CXN</sub>	F <sub>CRP</sub>
IMP	0.797	1.255	31.7%	6.1%
WHY	0.433	1.114	1.2%	0.1%
LETS	0.776	0.775	4.0%	0.9%
REQ	0.505	0.465	16.6%	10.0%
SEMI	0.176	0.302	13.5%	6.8%
TO.COMP	−0.056	−1.104	5.9%	17.1%
IND	−0.648	−1.778	4.5%	27.1%
Correlation with LMER		<b><i>r</i> = 0.93</b>	<b><i>r</i> = 0.31</b>	<b><i>r</i> = −0.88</b>
coefficients		<i>t</i> = 5.7, <i>p</i> < 0.01	<i>t</i> = 1.1, <i>p</i> = 0.35	<i>t</i> = −4.2, <i>p</i> < 0.01

the correlation with frequencies in the construction ( $F_{\text{CXN}}$ ), expressed as an average over *go/come*-V, is moderate, but not significant. Also recall that there was no effect of construction type (CXN; *go, come*) on acceptability (Section 4.4), despite the fact that *go*-V (13,049) is over three times more frequent than *come*-V (3528). Note at this juncture that the compatibility view accounts for *ceiling* and *floor* effects in the frequency/acceptability mismatch: WHY and LETS are as acceptable as IMP although they diverge very much in frequency (cf. *ceiling mismatch*), while LETS and IND, which are similar at least in  $F_{\text{CXN}}$ , diverge in acceptability (cf. *floor mismatch*).

A final point concerns the stability of distributional measures. The CA above is based on COCA, a balanced reference corpus with five broad genres (academic, fiction, magazine, news, spoken). Reference corpora come with the drawback that they are biased towards learned and mostly written material. They are thus highly unrepresentative of the language spoken or experienced by speakers in a speech community, especially of speakers with rare or no exposure to learned material. Spoken genres in a reference corpus or specialized corpora are not immune against this problem. For instance, the spoken section in COCA contains the language of TV debates from selected US networks, which is very different to the language of telephone conversations between strangers in the SWITCHBOARD corpus. In turn, SWITCHBOARD data is not typical of face-to-face communication. This textual variability adds to the problems when measuring (raw) frequency.

However, it appears that the correlation between judgment data and (complex) corpus distribution is robust across different data types. Table 5 shows the correlation for CA<sub>SRC</sub> and frequency measures from individual COCA genres and selected specialized corpora, for which separate CAs were calculated by the same procedure as described in Section 3.2. The slightly lower correlation coefficients for

the academic genre and CHILDES may reflect the fact that these data sources are rather atypical of adult speech.

The correlation between acceptability and  $CA_{SRC}$  is high and robust across corpora. This is, by and large, also true for the negative correlation with corpus frequencies ( $F_{CRP}$ ). The correlation with construction frequencies ( $F_{CXN}$ ) is lower, unsystematic, and statistically not significant. Whether this results from the small sample size or whether it indicates that construction frequencies are an unreliable measure across corpora cannot be read from this data.

However, there are reasons to assume that distributional measures are more robust than frequencies. How often a construction occurs (pmw,  $F_{CXN}$ ) in a corpus is context-dependent and thus sensitive to thematic or social variation. Hence, frequencies will vary considerably between corpora. What is crucial is that their *distributional* behavior remains stable. For instance, the strong correlation of acceptability with the conversational data in SWITCHBOARD appears unsurprising at first sight, given the informality of *go/come-v*. But since SWITCHBOARD contains telephone conversations between strangers, the imperative rate of *go/come-v* (9.1%) is much lower than in COCA (31.7%). In SWITCHBOARD, imperative *go/come-v* is even less frequent than indicative *go/come-v* (12.1%). But what is key is that directives and commissives are *overall* extremely low in SWITCHBOARD (IMP 1.5%) and the rate of indicatives is very high (40.8%; compared to COCA’s 27.1%). In other words, the skew of *go/come-v* towards directive contexts remains stable in comparison to average corpus samples, despite substantial differences in frequencies. Put simply, a CA bi-plot of

**Table 5:** Correlation between the LMER coefficients and the corpus measures by corpus type/genre; pmw is the per-million-word frequency of *go/come-v*.

CORPUS/GENRE <sup>14</sup>	pmw	$CA_{SRC}$	$F_{CXN}$	$F_{CRP}$
COCA, <i>full corpus</i>	34.9	0.93 **	0.31 ns	−0.88 **
Spoken	56.6	0.95 **	0.27 ns	−0.96 ***
Magazine	18.7	0.93 **	0.37 ns	−0.91 **
News	19.2	0.92 **	0.28 ns	−0.94 **
Fiction	76.8	0.89 **	0.38 ns	−0.96 ***
Academic	3.4	0.76 *	0.06 ns	−0.92 **
CHILDES (adult tokens only)	518.3	0.81 *	0.58 ns	−0.37 ns
ENCOW (web data)	19.9	0.93 **	0.49 ns	−0.88 **
SWITCHBOARD	184.2	0.91 **	−0.29 ns	−0.92 **

Note: SWITCHBOARD with fewer than 1000 observations per CXN due to corpus size.

**14** COCA (Davies 2008; late 2015 offline version), CHILDES (MacWhinney 2000; all xml-annotated files from North America), ENCOW (Schäfer and Bildhauer 2012; 2014 version, slice 03), SWITCHBOARD (Godfrey et al. 1992; *Open American National Corpus* version, [www.anc.org](http://www.anc.org)).

SWITCHBOARD data, representing the constructional ecology in telephone conversations, looks essentially identical to the COCA plot (Flach forthcoming). A similar argument holds in the other direction for CHILDES, where *go/come-v* is the most frequent in any of the corpora (518.3 pmw), and which has the highest rates of imperatives (40%) and the lowest rates of indicatives (2.3%). Yet, these do *not* automatically show the strongest correlation with judgment data (cf. Table 5), because directives and commissives are *generally* high in child-directed speech, i.e., *go/come-v* is less distinctive in child-directed speech.

In summary, there is a robust correlation between two types of performance data – corpus distribution as a proxy to speakers' experience with a construction and acceptability as a proxy to speakers' knowledge thereof. While the frequency of a unit or the distribution *within* that unit varies considerably between corpora and/or registers, distributional measures, determined in relation to other elements *across* a corpus, are relatively stable between data types. In other words, schema compatibility or a construction's ecology may be much less sensitive to imbalances or fluctuations in a specific data type.

## 6 General discussion

This article addressed the frequency/acceptability mismatch from the perspective of the schema, a central concept in cognitive usage-based models of language. Schemas are rich conceptual knowledge structures, which speakers extract from repeated exposure to instances of the same construction in their communicative habitat. From this angle, acceptability is a function of compatibility with a licensing schema, which accounts for the acceptability even of rare or corpus-absent patterns. While acceptability may be independent of (raw) frequency, it is not independent of usage intensity (FREQUENCY OF USAGE). The claim is that the frequency/acceptability mismatch arises in large parts from a methodological mismatch that tries to map simple measures onto complex syntactic phenomena.

Schema compatibility and its interplay with various usage-derived measures in the context of the frequency/acceptability mismatch was illustrated using the English *go/come-v* construction. A morphological constraint (BSC) leads to (un)grammaticality based on the presence or absence of inflection. A usage-based model explains the morphological constraint as the result of a semantic constraint: non-assertive constructional semantics make *go/come-v* functionally incompatible with inflectional contexts. The results of an acceptability judgment task with sentences that were specifically manipulated to reflect constructional semantics

confirmed that increasingly stronger violations of schema specifications correspond with decreased acceptability. The systematicity in gradient acceptability is difficult to account for if the BSC is seen as primarily morphological.

The study illustrates that the acceptability patterns are better captured by complex than by simplistic measures. In line with recent research, the frequency/acceptability mismatch is significantly reduced with distributional measures that represent a construction's usage properties more appropriately (Divjak 2017). Multidimensional measurements may also be considerably more robust across corpora or/and registers, which somewhat balances the extreme noisiness of corpora.

The experimental results indicate that acceptability is related to compatibility in two ways. The main influence is compatibility with the licensing schema. A minor influence emerges with respect to compatibility on a lower level: verb association positively affects acceptability, but only in cases of diminishing compatibility with the higher-order schema. Verb association, which can be seen as a tighter connection between two simpler constructions, may provide a fallback strategy that somewhat "saves" an otherwise semantically awkward structure. Put differently, non-association adds to soft constraint accumulation that affects acceptability (Langacker 2000; Sorace and Keller 2005).

It must be kept in mind that schema compatibility is a conceptual notion that may be very difficult to determine for structural units with higher type frequencies and/or multi-faceted interactions with smaller units, especially in morphological contexts (Arppe and Järvikivi 2007; Bermel et al. 2018; Divjak 2017). The identification of schema compatibility for *go/come-v* is straightforward, as the type frequency of two means that *go/come-v* is low in schematicity. The semantic-pragmatic dimension was easy to capture in just one variable (SYN). Yet, the construction illustrates the basic problem of the methodological mismatch rather well: units of varying degrees of specificity and schematicity require different measures of FREQUENCY OF USE (Stefanowitsch and Flach 2016: 106). Now recall that lexical frequency is an accurate predictor of experimental behavior in reaction time experiments for units that are low in schematicity (Arnon and Snider 2010; van Heuven et al. 2014). This is because units of low schematicity, such as simplexes (e.g., *old*, *young*, or *time*) or fixed complexes (e.g., *how do you do*), instantiate their own schema directly and are by definition compatible with it. (Except perhaps if they are manipulated in experimental conditions by violating selectional restrictions; manipulation decreases compatibility, because it affects connections to other units in the network.) The methodological mismatch does not arise here, at least not as quickly.

Yet, as the schematicity of a unit increases, so does the empirical complexity of the relationship between the licensing schema and its instantiation(s). Multiple

interactions can be at work, such as type frequencies, distributional skews (or their absence), productivity of open slots, or interference from overlapping constructions and their interactions with lexical elements. Most of the complex morphological and syntactic units will be affected by a number of usage properties. With increased complexity, a certain degree of unexplained or “unexplainable” variance or divergence between observed and elicited data is not surprising (see, e.g., Arppe and Järviö 2007; Bermel et al. 2018). This is no doubt due in no small part to the reductive nature of corpora or the incomplete coverage of speakers’ experience.

In a similar vein, the preceding discussion does not imply that there are measures that are inherently well-equipped to capture multiple phenomena across all levels of schematicity and complexity. There is a lively debate on the predictive power of different association measures (e.g., Gries 2015; Schmid and Küchenhoff 2013; Wiechmann 2008). However, it is doubtful whether it is possible or even necessary to identify (sets of) metrics that perform best across phenomena, experimental tasks, or types of questions. This is because compatibility with a schema depends on a number of very complex and interrelated factors that may be highly idiosyncratic to a given pattern. An anonymous reviewer remarks that this bears the danger of post-hoc curve fitting, making the theory unfalsifiable. However, the hypothesis was not that CA is a better predictor of acceptability than raw frequencies (or any other measure, for that matter), but that *simple* measures are outperformed by *complex* measures for complex phenomena (which is falsifiable). CA was used here because it is an appropriate method to represent the relationship between two categorical variables. The results neither imply that CA is suitable for (all) other phenomena, nor that CA or other contingent measure remove the frequency/acceptability mismatch entirely. But what this article *does* argue is that the role of experience is underestimated, if experience is approximated by unsuitable means.

An analogy might illustrate this point. A meteorological model that predicts the weather in the Alps will be unsuitable to predict the weather at the coast, because it ignores sea water evaporation and salinity, which are less relevant in the mountains. More generally, any model that considers complex ecological conditions will be better at predicting local weather than reductive, simple models – but it will trivially never be perfect or transfer straightforwardly to different ecologies. The point in the current case is that simple frequency counts are unsuitable because they miss important local properties; but there will always be some variation that cannot be accounted for. The goal is to tease apart which forces are at work in which area of the constructional space under which conditions. The effect of  $V_2$  association on semantic constraint violators is a result that would likely not have been detected in a (raw) frequency perspective.

Many studies which identify complex measures as better predictors of acceptability approach compatibility in a similar fashion. For instance, Gries et al. (2005)

use collostructional methods to predict sentence completion in the *as*-predicative, where verb association (*regard as*) reflects compatibility better than construction frequency (*see as*). Divjak (2017) uses morphological transparency of low-frequency and potentially unknown verbs in Polish *that*-clauses. Transparency refers to the degree to which a verb is recognizably related to – and hence more compatible with – typical verbs in the pattern. An approach not unlike the current one is Dąbrowska (2008), who investigates the acceptability of context manipulation in WH-questions with long-distance dependencies (*What do you think you are doing?*). The prototypical template that speakers are exposed to in everyday language contains a *do*-auxiliary, a pronoun subject (*you*), and a perception verb (overwhelmingly *think* or *say*). Although Dąbrowska's context manipulations are fully grammatical on a binary view, much like *go/come-v*, prototypical sentences received much higher ratings than progressive manipulations in the subject, verb, or auxiliary positions. Sentences received the lowest ratings if *all* of the slots deviated from the prototype, which is a form of maximal schema distance.

Now, whether gradient acceptability and/or the convergence of two types of performance data is viewed as theoretically relevant or extragrammatical will vary with one's model of language (Sprouse et al. 2018). Depending on what one means by "grammar", gradience will always be external to grammar (Newmeyer 2003), while some assume gradient grammaticality (Featherston 2005; Keller 2000; Wasow 2009; Weskott and Fanselow 2011). For usage-based theories, the competence–performance correspondence is more relevant, if not foundational. It crucially does not require the assumption of separate modular processes below a frequency threshold (Bader and Häussler 2010). But whether or not gradience is an integral part of one's linguistic model of competence, the results in this study suggest that the correspondence between usage (or corpus) distribution and experimental behavior may have methodological implications. That is, complex usage properties beyond lexical or structural frequencies may need to be factored into experimental items and analyses, even if only to remove a performance effect that is irrelevant for the competence or knowledge one is interested in.

**Acknowledgments:** The research reported in this article was partly funded by the Swiss National Science Foundation, SNF Grant 100012L/169490/1; support is gratefully acknowledged. I thank Martin Hilpert, Kristin Kopf, and Karin Madlener for discussion and valuable comments, as well as three anonymous reviewers for their constructive feedback and the food for thought. The usual disclaimers apply.



# Appendix

## Corpus data

Extracted from COCA (Davies 2008): case-insensitive strings <go> and <come> followed by verbs or nouns (e.g., *Let's go party*<sub><N></sub>), then manually cleaned (cf. Flach 2017b). **IMP**: instances in the imperative (*Go see a doctor*, *Somebody come rescue me!*). **LETS**: *let's* and *let us* with speaker–hearer inclusive *us* (*Let's go have a drink*; *Let us go have a drink*). **WHY**: present tense uses of *why don't NP* (*Why don't you come visit?*). **MOD**: core modal auxiliaries (*can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, *would*) and modal idioms (*would rather* and (*had*) *better*), including interrogatives (*Should he go-v?*). **SEMI**: semi-modals (*have [got] to*, *need to*, *want to*, *going to*, *ought to*, *used to*, *dare [to] go/come-v*). **REQ**: bi-clausal patterns with a requestive matrix verb (e.g., *ask/force/order/invite/require NP to go/come-v*), subjunctives (*recommend/suggest [that] NP*), and adjectival patterns (*be supposed to*, *be ready to*, *be welcome to*). **TO.COMP**: subordinating patterns (*it is time to go/come-v*, *they saw him go-v*) and pseudo-clefts (*what she did was go-v*). **DO**: do-support (*Did he go eat?*, *he does/did go-v*). **IND**: indicative *go/come-v* (*I go get the mail*).

**Experiment items:** RK = rank of item in list (reverse for a list's alternative order).

SENTENCE	LIST	QID	RK	SYN	V <sub>2</sub>	ASSOC
Go find help immediately!	A	1	16	IMP	find	yes
Go seek help immediately!	B	2	28	IMP	seek	no
Go bring me some water, please.	A	3	27	IMP	bring	no
Go get me some water, please.	B	4	1	IMP	get	yes
Let's go make dinner together!	A	5	7	LETS	make	yes
Let's go cook dinner together!	B	6	25	LETS	cook	no
Let's go drink a cold beer.	A	7	15	LETS	drink	no
Let's go have a good burger.	B	8	30	LETS	have	yes
He was told to go wash his face.	A	9	30	REQ	wash	yes
He was told to go wipe his face.	B	10	26	REQ	wipe	no
Mum asked us to go defend our brother.	A	11	20	REQ	defend	no
Mum asked us to go help our brother.	B	12	18	REQ	help	yes
We need to go take this stuff outside.	A	13	4	SEMI	take	yes
We need to go bring this stuff outside.	B	14	8	SEMI	bring	no
I'm gonna go send them an email.	A	15	9	SEMI	send	no
I'm gonna go write them an email.	B	16	3	SEMI	write	yes
She likes to go play chess.	A	17	24	TO.COMP	play	yes

## Experiment items: (continued).

SENTENCE	LIST	QID	RK	SYN	V <sub>2</sub>	ASSOC
She likes to go learn languages.	B	18	13	TO.COMP	learn	no
It's time to go seek work.	A	19	21	TO.COMP	seek	no
It's time to go find work.	B	20	22	TO.COMP	find	yes
I go fetch the mail regularly.	A	21	29	IND	fetch	yes
I go retrieve the mail regularly.	B	22	11	IND	retrieve	no
They go relax on the couch.	A	23	18	IND	relax	no
They go sleep on the couch.	B	24	9	IND	sleep	yes
My sister went buy the groceries.	A	25	6	PST	buy	yes
My sister went bought the groceries.	B	26	15	PST	buy	yes
We had gone watched a movie.	A	27	12	PRT	watch	yes
We had gone look at the mess.	B	28	16	PRT	look	yes
Jake goes talks nonsense.	A	29	2	3RD	talk	yes
Jake goes talk nonsense.	B	30	6	3RD	talk	yes
Come visit the countryside this weekend!	A	31	22	IMP	visit	yes
Come explore the countryside this weekend!	B	32	17	IMP	explore	no
Come read me a story now!	A	33	5	IMP	read	no
Come tell me a story now!	B	34	10	IMP	tell	yes
Why don't you come see me after class?	A	35	1	WHY	see	yes
Why don't you come ask me after class?	B	36	4	WHY	ask	no
Why don't you come have lunch with me?	A	37	25	WHY	have	no
Why don't you come eat lunch with me?	B	38	14	WHY	eat	yes
Dad invited them to come stay with us.	A	39	10	REQ	stay	yes
Dad invited them to come live with us.	B	40	7	REQ	live	no
Dan called someone to come repair the sink.	A	41	3	REQ	repair	no
Dan called someone to come fix the sink.	B	42	12	REQ	fix	yes
You have to come get me out.	A	43	17	SEMI	get	yes
You have to come bail me out.	B	44	21	SEMI	bail	no
Lisa's gonna come teach at our department.	A	45	26	SEMI	teach	no
Lisa's gonna come work at our department.	B	46	20	SEMI	work	yes
It's a chance to come help a friend.	A	47	13	TO.COMP	help	yes
It's a chance to come support a friend.	B	48	2	TO.COMP	support	no
Children love to come hear stories.	A	49	8	TO.COMP	hear	no
Children love to come watch movies.	B	50	29	TO.COMP	watch	yes
They come fix computers for a living.	A	51	23	IND	fix	yes
They come repair computers for a living.	B	52	19	IND	repair	no
Our parents come travel with us every summer.	A	53	14	IND	travel	no
Our parents come stay with us every summer.	B	54	23	IND	stay	yes
He came picked me up from school.	A	55	19	PST	pick	yes
He came pick me up from school.	B	56	5	PST	pick	yes
She has come meet me often.	A	57	28	PRT	meet	yes
She has come met me often.	B	58	24	PRT	meet	yes
Helen comes speak to her parents daily.	A	59	11	3RD	speak	yes
Helen comes speaks to her parents daily.	B	60	27	3RD	speak	yes

Detailed descriptive statistics

Table 6: Detailed mean ratings by construction type (*go*, *come*) and V<sub>2</sub> association (*yes*, *no*).

SYN	go-v		come-v	
	yes	no	yes	no
<b>Associated V<sub>2</sub>:</b>				
IMP	0.84	0.66	0.65	0.42
WHY	–	–	0.72	0.57
LETS	0.52	0.77	–	–
REQ	0.62	0.46	0.70	0.71
SEMI	0.57	0.01	0.42	0.35
TO.COMP	0.44	–0.36	0.09	0.33
IND	–0.36	–0.50	–0.01	–0.65
<b>Inflected V<sub>2</sub>:</b>				
	yes	no	yes	no
PST	–1.23	–0.87	–1.29	–1.33
PRT	–1.41	–1.52	–1.21	–1.40
3RD	–1.51	–1.75	–1.05	–1.44

Note: LETS and WHY were restricted to *go* and *come* items, respectively. The inflected contexts PST, PRT, and 3RD were only paired with associated V<sub>2</sub>; hence, their mean ratings are distinguished of whether V2 was inflected.

References

Ambridge, Ben & Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19(3). 357–389.

Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1). 67–82.

Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.

Baayen, R. Harald, Doug J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Baayen, R. Harald & Dagmar Divjak. 2017. Ordinal GAMMs: A new window on human ratings. In Anastasia Makarova, Stephen M. Dickey & Dagmar Divjak (eds.), *Each venture a new beginning: Studies in honor of Laura A. Janda*, 39–56. Bloomington, Indiana: Slavica.

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330.

Bermel, Neil & Luděk Knittl. 2012. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory* 8(2). 241–275.

- Bermel, Neil, Luděk Knittl & Jean Russell. 2018. Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. *Corpus Linguistics and Linguistic Theory* 14(2). 197–231.
- Bjorkman, Bronwyn M. 2016. *Go get, come see*: Motion verbs, morphological restrictions, and syncretism. *Natural Language & Linguistic Theory* 34(1). 53–91.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, CA: Sage Publications.
- Crocker, Matthew W. & Frank Keller. 2006. Probabilistic grammars as models of gradience in language processing. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199274796.001.0001>.
- Dąbrowska, Ewa. 2008. Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics* 19(3). 391–425.
- Davies, Eirlys. 1986. *The English imperative*. London: Croom Helm.
- Davies, Mark. 2008. The corpus of contemporary American English: 450 million words, 1990–present. (2015 Offline Version). Available at: [corpora.byu.edu/coca](http://corpora.byu.edu/coca).
- Divjak, Dagmar. 2008. On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools, and applications – state of the art*, 213–233. Frankfurt: Peter Lang.
- Divjak, Dagmar. 2017. The role of lexical frequency in the acceptability of syntactic variants: Evidence from *that*-clauses in Polish. *Cognitive Science* 41(2). 354–382.
- Endresen, Anna & Laura A. Janda. 2017. Five statistical models for Likert-type experimental data on acceptability judgments. *Journal of Research Design and Statistics in Linguistics and Communication Science* 3(2). 217–250.
- Featherston, Sam. 2005. The decathlon model of empirical syntax. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 187–208. Berlin: Mouton de Gruyter.
- Firth, John R. 1957. A synopsis of linguistic theory, 1930–55. In Frank R. Palmer (ed.), *Selected papers of John Rupert Firth (1952–59)* (Published 1968), 168–205. London: Longmans.
- Flach, Susanne. 2015. Let's go look at usage: A constructional approach to formal constraints on *go-VERB*. In Peter Uhrig & Thomas Herbst (eds.), *Yearbook of the German Cognitive Linguistics Association*, vol. 3, 231–252. Berlin: De Gruyter Mouton.
- Flach, Susanne. 2017a. *Collostructions: An R implementation for the family of collostructional methods*. Available at: <https://sfla.ch/collostructions/>.
- Flach, Susanne. 2017b. *Serial verb constructions in English: A usage-based approach*: Freie Universität Berlin PhD thesis.
- Flach, Susanne. forthcoming. *Formal constraints in a usage-based perspective*. Berlin: De Gruyter Mouton.
- Glynn, Dylan. 2014. Correspondence analysis: Exploring data and identifying patterns. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 443–485. Amsterdam: John Benjamins.
- Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE international conference on acoustics, speech, and signal processing (ICASSP-92)*, 517–520. <https://doi.org/10.1109/ICASSP.1992.225858>. Available at: [www.anc.org](http://www.anc.org).

- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Greenacre, Michael J. 2017. *Correspondence Analysis in practice*, 3rd edn. Boca Raton: Chapman & Hall/CRC.
- Gries, Stefan Th. 2015. More (old and new) misunderstandings of collocation analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4). 635–676.
- Harris, Zellig S. 1954. Distributional structure. *Word* 10(2–3). 146–162.
- Häussler, Jana & Tom S. Juzek. 2016. Detecting and discouraging non-cooperative behavior in online experiments using an acceptability judgment task. In Hanna Christ, Daniel Klenovšák, Lukas Sönning & Valentin Werner (eds.), *A blend of MaLT: Selected contributions from the methods and linguistic theories symposium 2015*, 73–99. Bamberg: University of Bamberg Press.
- Jaeggli, Osvaldo A. & Nina M. Hyams. 1993. On the independence and interdependence of syntactic and morphological properties: English aspectual *come* and *go*. *Natural Language & Linguistic Theory* 11(2). 313–346.
- Keller, Frank. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Edinburgh: University of Edinburgh PhD thesis. Available at: <http://homepages.inf.ed.ac.uk/keller/publications/phd.pdf>.
- Kempen, Gerard & Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Stephan Kepser & Marga Reis (eds.), *Linguistic evidence*, vol. 85, 329–350. Berlin: Mouton de Gruyter.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, Ronald W. 1988. A usage-based model. In Brygida Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*, 127–161. Amsterdam: John Benjamins.
- Langacker, Ronald W. 2000. A dynamic usage-based model. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 24–63. Stanford, CA: CSLI Publications.
- Lau, Jey Han, Alexander Clark & Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* 41(5). 1202–1241.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum.
- Manning, Christopher D. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 289–341. Cambridge, MA: MIT Press.
- Nenadić, Oleg & Michael Greenacre. 2007. Correspondence Analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software* 20(3). 1–13.
- Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79(4). 682–707.
- Palan, Stefan & Christian Schitter. 2018. Prolific.ac: A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17. 22–27.
- Pullum, Geoffrey K. 1990. Constraints on intransitive quasi-serial verb constructions in modern colloquial English. In Brian D. Joseph & Arnold M. Zwicky (eds.), *When verbs collide: Papers*

- from the 1990 Ohio State mini-conference on serial verbs, 218–239. Columbus, OH: Ohio State University.
- Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, 486–493. Istanbul: ELRA.
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101–133. Berlin: De Gruyter Mouton.
- Schmid, Hans-Jörg. 2013. Is usage more than usage after all? The case of English *not that*. *Linguistics* 51(1). 75–116.
- Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577.
- Schütze, Carson T. & Jon Sprouse. 2013. Judgement data. In Robert Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, 27–50. Cambridge: Cambridge University Press.
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society* 5(1). 1–23.
- Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115(11). 1497–1524.
- Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001–2010. *Lingua* 134. 219–248.
- Sprouse, Jon, Beracah Yankama, Sagar Indurkha, Sandiway Fong & Robert C. Berwick. 2018. Colorless green ideas do sleep furiously: Gradient acceptability and the nature of the grammar. *The Linguistic Review* 35(3). 575–599.
- Stefanowitsch, Anatol. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3). 513–531.
- Stefanowitsch, Anatol & Susanne Flach. 2016. The corpus-based perspective on entrenchment. In Hans-Jörg Schmid (ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, 101–127. Berlin: De Gruyter.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Taylor, John R. 2004. The ecology of constructions. In Günter Radden & Klaus-Uwe Panther (eds.), *Studies in linguistic motivation*, 49–73. Berlin: Mouton de Gruyter.
- van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67(6). 1176–1190.
- Wasow, Thomas. 2009. Gradient data and gradient grammars. *Chicago Linguistics Society* 43. 255–271.
- Weskott, Thomas & Gisbert Fanselow. 2011. On the informativity of different measures of linguistic acceptability. *Language* 87(2). 249–273.
- Wiechmann, Daniel. 2008. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2). 253–290.